END

DATE
FILMED

8 8!

DTIC

AD A101992

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ONR 81-03	2. GOVT ACCESSION NO. AD-A101 992	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Comparison of two treatments when there may be an initial effect		5. TYPE OF REPORT & PERIOD COVERED Scientific
		6. PERFORMING ORG. REPORT NUMBER ONR 81-03
7. AUTHOR(s) Professor Elizabeth L. Scott		8. CONTRACT OR GRANT NUMBER(s) ONR N00014 75 C 0159
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistical Laboratory Department of Statistics University of California, Berkeley CA 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington DC 20014		12. REPORT DATE June 1981
		13. NUMBER OF PAGES 20 pages
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

treatments with two mechanisms
Neyman $C(\alpha)$ tests
power function
gamma distribution
multiplicative effect

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Consider the situations where the treatment may cause an initial effect and may also cause a long-range effect. We want to evaluate the treatment, or to compare two treatments, when the effect of treatment may result from the two distinct mechanisms, M_1 and M_2. We may wish to evaluate M_1 and M_2 separately, but we may also want to evaluate their combined effect M_3.

Examples are given and the general rules are applied to the special case

(continued on reverse)

DD FORM 1473   EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73

Accession For

NTIS GRA&I
DTIC TAB
Unannounced
Justification

By
Distribution/
Availability Codes

	Avail and/or
Dist	Special

A

20. (continued)

arising in weather modification studies and elsewhere: the possible effects are
multiplicative and the distribution of nonzero variables is Gamma with at most
the scale parameter affected by treatment. An example demonstrates that the
two components may be too weak to be judged significant while their sum is
large and significant. The locally optimum $C(\alpha)$ test is used.

There is a brief discussion of the power function of the tests.

COMPARISON OF TWO TREATMENTS WHEN THERE MAY BE AN INITIAL EFFECT

ELIZABETH L. SCOTT, University of California, Berkeley

Abstract

Consider situations where the treatment may cause an initial effect and may also cause a long-range effect. We want to evaluate the treatment, or to compare two treatments, when the effect of treatment may result from the two distinct mechanisms, M_1 and M_2. We may wish to evaluate M_1 and M_2 separately, but we may also want to evaluate their combined effect M_3. Examples are given and the general results are applied to the special case arising in weather modification studies and elsewhere: the possible effects are multiplicative and the distribution of nonzero variables is Gamma with at most the scale parameter affected by treatment. An example demonstrates that the two components may be too weak to be judged significant while their sum is large and significant. The locally optimum $C(\alpha)$ test is used.

There is a brief discussion of the power function of the tests. The asymptotic power agrees well, in general, with the results of the Monte Carlo simulation for the test Z_3 of the combined effect. If the zero values are discarded and then Z_2 employed, there is large bias in the power. The bias is more pronounced if the Wilcoxon, Mann-Whitney test is employed. Notice that the two effects under study may be acting in the same direction or they may be in opposition.

TREATMENTS WITH TWO MECHANISMS, NEYMAN $C(\alpha)$ TESTS, POWER FUNCTION, GAMMA DISTRIBUTION, MULTIPLICATIVE EFFECT

1. Introduction

We consider situations where the treatment may cause an initial effect and may also cause a long-range effect. There are many examples. Mosteller (1977) described the Portacaval Shunt Operation "designed to reduce pressure from the blood stream in the esophagus and thus prevent or stop hemorrhaging in the patient. The operation has a substantial death rate". Other treatments and also nontreatment have a substantial death rate. The Portacaval Shunt Operation may (1) affect the probability of surviving the initial period of treatment and also may (2) affect the number of years of survival of those patients who do live through the operation. The two effects may be in the same direction or in opposite directions. Although both effects are of concern to the patient and his physician, the combined effect is also important.

We want to evaluate the treatment, or to compare two treatments, when the effect of treatment may result from two distinct mechanisms, denoted by M_1 and M_2, say. Mechanism M_1 consists in the possible modification of the probability of an initial effect of treatment. The hypothesis that no such effect occurs will be denoted by H_1. Then, mechanism M_2 consists in a change in the conditional distribution of the variable under study, say Y, given that $Y > 0$. The hypothesis that M_2 is not operating will be denoted by H_2. We may wish to evaluate M_1 and M_2 separately, but we may also want to estimate their combined effect M_3, the total change per experimental unit. The distinction between the mechanisms M_1 and M_2 and their combined effect M_3 is often ignored. This may be unfortunate since the separate effects of the initial mechanism and of

the long-range mechanism may be weak and therefore difficult to detect while their combined effect may be important and capable of detection. On the other hand, M_1 and M_2 may be in opposition so that they tend to cancel each other. In this case, an analysis of either alone may be misleading.

A second example where the treatment may act through two mechanisms arises in the treatment of cancer (and other diseases). M_1 could alter the probability of unpleasant side effects that could force the patient to withdraw from treatment and/or be fatal. M_2 may affect the conditional expected length of survival, given that the patient continues treatment. As an illustration, for some diagnoses of cancer, the standard treatment is a harsh chemical program which some patients cannot withstand. A new treatment consists of the administration of a transfer factor designed to increase the patient's immunity to his/her specific kind of cancer. The statistician consulted on such an experiment may want to compare the treatments by comparing the performance of all patients assigned to one treatment with the performance of all patients assigned to the other. However, the physician may feel that those patients who withdrew from treatment early in the experiment, for whateever reason, have been administered so little treatment that their inclusion would not be meaningful and would tend to dilute the results. Actually, the statistician wants to study mechanism M_3 and the physician wants to study M_2.

In many examples, the distribution of survival time is nonstandard. A further complication arises when the experimental units are not homogeneous. In the example above, the patients may differ with respect to age, sex,

level of diagnosis, and so forth. These characteristics of the unit may serve as predictor variables for the experimental variable under study.

Speaking generally, we consider a randomized experiment with independent units. For the k-th unit, let

X_k = the predictor variable with probability density $f(x_k;\lambda)$ where

X and λ may be vectors;

T_k = 1 if the new treatment is applied, 0 if not, with $Pr\{T=1\} = \pi$;

Y_k = the experimental variable with probability density $p[y|x,\theta(t,\xi)]$

of known form, vector parameters.

We assume that the effect of treatment enters through ξ, as follows. If $\theta_j(t,\xi) = \theta_j$ when $T = 0$, then when $T = 1$ with the same value x, we have

$$(1.1) \quad \theta_j(t,\xi) = \theta_j + \theta_j'\xi + o(\xi), \qquad \text{for } j = 1, \ldots, s.$$

We thus have a triplet (X,T,Y) for each experimental unit, with probability density, say,

$$(1.2) \quad \psi(x,t,y) = \pi^t(1-\pi)^{1-t} f(x;\lambda) p[y|x,\theta(t,\xi)].$$

The hypothesis of no effect becomes the hypothesis $\xi = 0$. Neyman and Scott (1967) have found the locally optimum test of class $C(\alpha)$ to have as test criterion

$$(1.3) \quad Z = \frac{\sum_{k=1}^{n} (t_k-\pi) \sum_{j=1}^{s} \theta_j' \phi_j(y_k,x_k,\hat{\theta})}{n^{\frac{1}{2}}\{\pi(1-\pi) \text{Var}[\sum_{j=1}^{s} \theta_j' \phi_j(y_k,x_k,\hat{\theta})]\}^{\frac{1}{2}}}$$

where

$$\phi_j = \frac{\partial \log p(y|x,\theta)}{\partial \theta_j}\Bigg|_{\xi=0} \qquad\qquad j = 1, \ldots, s.$$

The criterion Z is asymptotically Normal$(0,1)$ when $\xi = 0$, and is noncentral Normal when $\xi \neq 0$, with noncentrality parameter equal to ξ multiplied by the denominator of Z.

Neyman and Scott (1965, 1967) noted that the test criterion (1.3) does not depend on the distribution f of the predictor variable X except that the denominator must take into account the variability of the predictor X as well as that of the experimental variable Y. Moran (1973) extended this result.

In the situation of this paper, we may wish to consider three problems separately:

1) We may want to estimate ξ_1, the effect of M_1, or we may want to test that M_1 has no effect which would correspond to $\xi_1 = 0$,

2) We may want to estimate ξ_2, the effect of M_2, or we may want to test that M_2 has no effect which would correspond to $\xi_2 = 0$,

3) We may want to estimate the combined effect ξ_3, or we may want to test that M_3 has no effect which would correspond to $\xi_3 = 0$.

We thus employ (1.3) to develop three test criteria Z_1, Z_2, and Z_3. A case of wide application is considered in the next section.

2. Case of multiplicative effect accompanied by a Gamma distribution

In many applications, we can assume that if the effect occurs at all, it is multiplicative. Often, we can assume that the distribution of the nonzero variable is Gamma with shape parameter unaffected by the treatment. We then have for the initial mechanism M_1

$$\theta(\xi_1) = \theta(1 + \xi_1)$$

so that ξ_1 measures the proportional improvement,

$$(2.1) \qquad \xi_1 = [\theta(\xi_1) - \theta]/\theta.$$

For mechanism M_2, on combining the assumption that the nonzero effect is multiplicative with the assumption of a Gamma distribution with at most the scale parameter affected, we have

$$(2.2) \qquad p_Y(y|\gamma,\delta) = \frac{\delta^\gamma}{\Gamma(\gamma)} \, y^{\gamma-1} \, e^{-\delta y},$$

where $\gamma > 0$ is the shape parameter and $\delta > 0$ is the inverse of the scale parameter. Under the assumption that treatment can affect only the scale parameter,

$$(2.3) \qquad \delta_t = \delta(\xi_2) = \frac{\delta_u}{1 + \xi_2} , \qquad \begin{cases} t = \text{treated (new treatment)}, \\ u = \text{untreated (standard)}. \end{cases}$$

The analysis of weather modification experiments is an example where the assumptions of multiplicative effect and of a Gamma distribution of the nonzero effects are well satisfied. In fact, this application led to

the development of the problem (Neyman and Scott, 1967). Earlier analyses used designs with comparison areas under Normal theory (cf. Moran, 1955) and then under locally optimal $C(\alpha)$ theory (Neyman, Scott, and Vasilevskis, 1960). Moran showed that the comparison area design with cross-over of treatment is advantageous when applicable. However, the effects of weather modification appear to be widespread causing contamination of the comparison area. Similar difficulties can arise in other types of application; we will restrict attention to randomized trials on homogeneous units.

Cloud seeding, for example by the release of silver iodide into clouds in an effort to provide nuclei for the condensation of water vapor, could possibly cause precipitation to reach the ground which would not have fallen otherwise (or conversely). In addition, the cloud seeding may increase (or decrease) the amount of precipitation falling, given that there is some precipitation. Thus, we have a mechanism M_1 and a mechanism M_2 that may be acting in the same or in opposite directions. The total effect depends on the combination of the two mechanisms.

Meteorologists predict that both of the postulated mechanisms will be multiplicative. The distribution of nonzero precipitation is typically a Gamma distribution, and as illustrated in Figure 1, this approximation is reasonable even when the same shape parameter is employed for both the seeded and the not-seeded experimental units, at lease for similar types of storms. When the storm categories may differ, it is reasonable (Dawkins, Neyman, Scott, and Wells, 1977) to assume that the experimenters can predict the category before treatment starts, and before the randomized decision to treat or not treat the storm is made. For example, the experimenters can predict the duration $D = d_k$ for the k-th experimental unit, and can assume that its effect
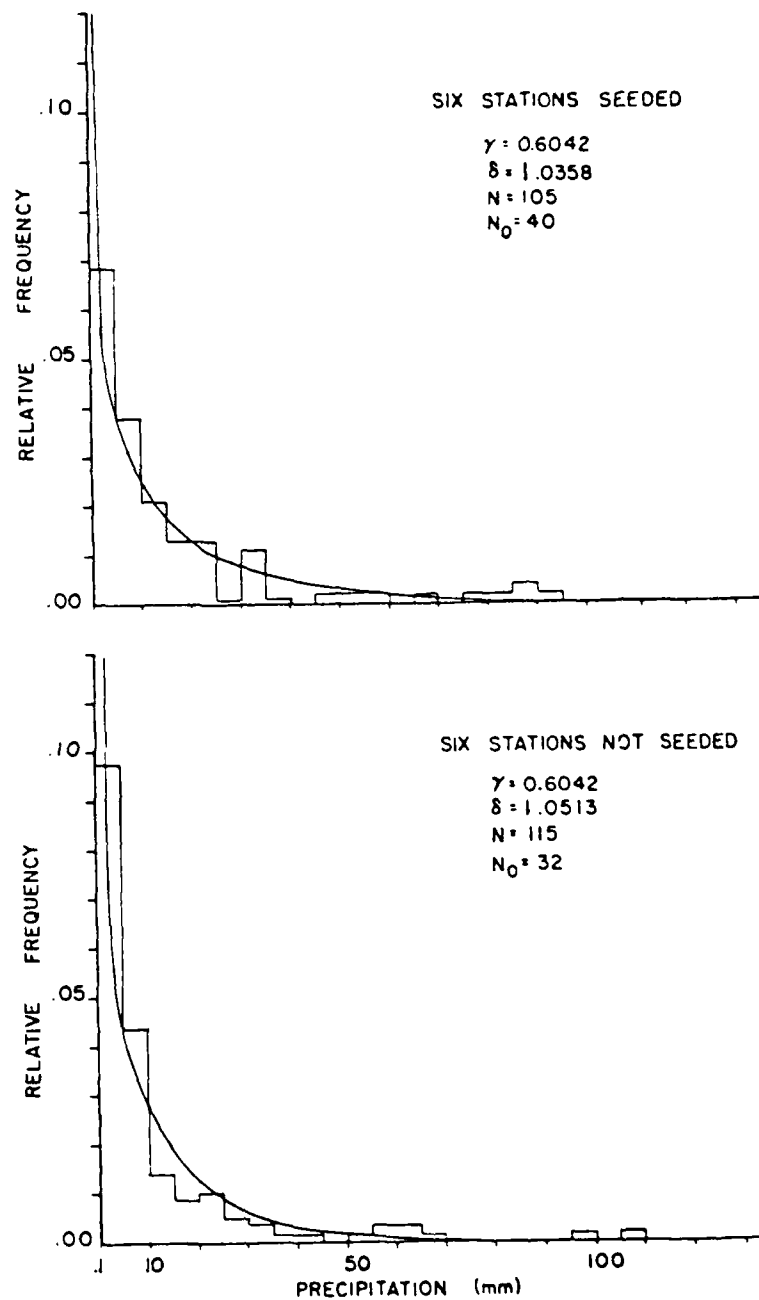
Figure 1

Typical comparison of observed distribution of nonzero precipitation
with Gamma distribution fitted by maximum likelihood with same shape
parameter. These data correspond to the six stations with altitude
< 1000 km in zone 4 of the Swiss hail experiment Grossversuch III.

enters the distribution of untreated precipitation only through the shape parameter which can be approximated linearly,

$$\gamma(d) = A_0 + A_1(d - d_0) \ .$$

Here, d_0 is the population mean of the variable predicted duration, and A_0, A_1, and δ_u are unknown 'nuisance parameters'. Notice that the differences in storm types, which may be complex, have a summary effect on the distribution of precipitation which may be quite large but can be summarized by changes in the shape of the distribution of nonzero precipitation expressed as a linear function of the predicted duration. Since the effect of seeding, if any, is assumed to alter only the scale parameter, we have that the conditional distribution of nonzero precipitation, given the predicted duration, is a Gamma distribution with constant shape parameter.

Our experience indicates that similar assumptions can be made in other fields of application, for example in survival analysis for clinical trials.

Under the assumptions and notation adopted, the expected value of the precipitation in a treated unit is

$$(2.4) \qquad E(Y_t) = \theta(\xi_1) \ A_0 \ / \ \delta(\xi_2) = A_0 \ \theta \ (1+\xi_1)(1+\xi_2) \ / \ \delta_u.$$

For a fixed value of D, that is, for a fixed category of storm types, the expected percent effect of seeding, due to both mechanisms, is

$$(2.5) \quad \text{Percent effect} = 100 \ [(1+\xi_1)(1+\xi_2) - 1] = 100(\xi_1 + \xi_2 + \xi_1\xi_2).$$

The theory of $C(\alpha)$ tests refers to certain limiting situations where the number of observations is "large" and the effects of treatment, such as ξ_1 and ξ_2, are "small". In consequence, we have tended to adjust the test of M_3 to be particularly sensitive to

$$\xi_1 + \xi_2 = \eta, \text{ say,}$$

neglecting the product term $\xi_1\xi_2$. In typical weather modification experimentation (also in clinical trials), ξ_1 might be 0.1 and ξ_2 might be 0.2 so that the neglected product is only 0.02.

3. Application

The test criterion for the individual tests are found to be (Neyman and Scott, 1967a), using $C(\alpha)$ tests:

For the hypothesis H_1 that $\xi_1 = 0$, which means that the probability of _initial_ effect (the probability of initiating precipitation) is not altered by treatment, corresponds to the familiar chi for this case:

Number of Initial Reactions

	Treated	Untreated	Total
React	n_{+t}	n_{+u}	$n_{+.}$
Do not react	n_{0t}	n_{0u}	$n_{0.}$
Total			n

(3.1)　　　$Z_1 = [n_{+t}\, n_{0u} - n_{+u}\, n_{0t}] \,/\, [n\pi(1-\pi)\, n_{+.}\, n_{0.}]^{\frac{1}{2}}$,

where the notation is set out in the usual table and π is the
adopted probability of treatment. The significance probability of
Z_1 is the two-tailed Normal probability; reject H_1 if $|Z_1| \geqq \nu(\alpha)$,
the critical value corresponding to level α.

For the hypothesis H_2 that $\xi_2 = 0$, which means that, given
that the initial effect is survival (given that there is nonzero
precipitation), there is no effect of treatment (the scale parameter
in the Gamma distribution of nonzero precipitation is not altered),
the test criterion turns out to be (Dawkins, Neyman, Scott, and Wells, 1977)

(3.2)　　　$Z_2 = \{\sum_t [\hat{\delta} y_k - \hat{\delta}(d_k)] - \sum_u [\hat{\delta} y_k - \hat{\delta}(d_k)]\} \,/\, (n\hat{A}_0)^{\frac{1}{2}}$,

where the sums are taken over the $t =$ treated and $u =$ untreated
units separately. The $\hat{A}_0$, $\hat{A}_1$, and $\hat{\delta}$ are solutions of the

simultaneous maximum likelihood equations taken over all nonzero units, treated and untreated,

$$\sum \{\Psi\{\hat{\gamma}(d_k)\} - n \hat{A}_0 = \sum \log_e y_k - n \log_e(\sum y_k / n)\ ,$$

$$\sum(d_k - \bar{\bar{d}})\Psi\{\hat{\gamma}(d_k)\} = \sum(d_k - \bar{\bar{d}}) \log_e y_k,$$

$$\hat{\delta} \approx n \hat{A}_0 / \sum y_k,$$

with $\bar{\bar{d}} = \sum d_k / n$, the grand mean, and Ψ is the derivative of the Gamma function. Also,

$$\hat{\gamma}(d_k) = \hat{A}_0 + \hat{A}_1(d_k - \bar{\bar{d}}).$$

The significance probability of Z_2 is two-tailed Normal asymptotically.

For the hypothesis H_3 that the combined effect of treatment is zero, we have as noted above been testing that $\xi_1 + \xi_2 = 0$. The test criterion Z_3 is a weighted sum of Z_1 and Z_2,

$$(3.3) \quad Z_3 = (\Delta_1 Z_1 + \Delta_2 Z_2) / (\Delta_1^2 + \Delta_2^2)^{\frac{1}{2}}\ ,$$

with

$$\Delta_1^2 = \hat{\theta} / (1-\hat{\theta}) = n_{+.} / n_{0.}\ ,$$
$$\Delta_2^2 = A_0^{\star}\ ,$$

which is the solution of the system of maximum likelihood equations when both δ_t and δ_u are entered as separate and possibly different parameters. The significance probability of Z_3 is two-tailed Normal asymptotically.

The application of the three test criteria is illustrated in Table 1, referring to the evaluation of hail reports from the Grossversuch III hail suppression experiment in southern Switzerland (Sänger *et al*, 1958-64). An earlier analysis (Neyman, Scott, and Wells, 1966) of the effects of seeding on rainfall (which is easier to observe than hail) suggested that the effect of seeding is positive, with a significant increase in rainfall, when there are stability layers in the atmosphere, as indicated on the early morning nearest radiosonde observed at Milan. It is of interest, then, to study the effects of hail for this category of days. The results are shown in Table 1. The first rows of the table refer to the category of days 'without stability layers' first for seeded (S) and then for nonseeded (NS) days. The next two rows refer to days 'with stability layers', and the last two rows to all days combined. The first block of results refers to mechanism M_1 -- is the frequency of days with hail altered by seeding? There is an indication of an increase of +54% for the category of days with stability layers, but the increase is not significant by the usual standards; the two-tail significance probability corresponding to the test criterion Z_1 is only 0.093. *There is no suggestion of change on days without stability layers.*

When we examine the second mechanism M_2, we note that the amount of hail per day with hail appears to be increased by +47% but the effect is not significant, P now being 0.17 for the experimental days with stability layers on which there was hail, as estimated by the

TABLE 1

Evaluation of hail reports for Grossversuch III target data.

Effect of seeding on frequency of hail reports, on number of

hail reports per hail day, and on number of hail reports

per experimental day.

CATEGORY OR DAYS	SEEDING	FREQUENCY				AMOUNT PER HAIL DAY			AMOUNT PER EXPERIMENTAL DAY		
		No.of zero	No.of hail.	Percent effect	P	Mean	Percent effect	P	Mean	Percent effect	P
Without stability layers	S	32	15	0	1.00	2.80	-23	0.40	0.89	-23	0.54
	NS	34	16			3.62			1.16		
Stability layers	S	63	33	+54	0.093	6.67	+47	0.17	2.29	+127	0.024
	NS	73	21			4.52			1.01		
All days	S	97	48	+32	0.17	5.46	+32	0.19	1.81	+74	0.049
	NS	110	37			4.14			1.04		

asymptotic criterion Z_2. When we continue to the totality of experimental days with stability layers and use Z_3 to evaluate the possible change in the number of hail reports per experimental day on which there are stability layers, we find an estimated increase of +127% with significance probability 0.024. Thus the combined effect of the two mechanisms is large and significant; they appear to be acting in the same direction.

However, on experimental days without stability layers, the estimated effect is a small decrease, but it is far from significant. On all experimental days whatsoever, the estimated increase is positive +74% and $P = 0.049$ significant at the standard level.

We thus have evidence that both mechanisms are playing a role: there is some evidence of an increase in the probability of hail and, given that there is hail reported, there is some evidence of an increase in the number of hail reports. As occurred when rainfall was the experimental variable, we find the positive effect is pronounced on the experimental days with stability layers. Since the purpose of the cloud seeding was to reduce hail , it appears that seeding with silver iodide is counter-indicated, at least as performed in this experiment, on days with stability layers. If the experiment is analysed using only days with positive hail reports -- comparing the hail counts on hail days that were seeded with those when there was no seeding but discarding the days with no hail reports (as is done with some operators) -- the estimated effects (as shown in the middle part of Table 1) would be much smaller, not significant, and possibly misleading.

4. Discussion of the power function

The probability of detecting an effect when it exists is asymptotically noncentral Normal for each of the three test criteria. The power function of Z_3, the criterion for combined effect of the two mechanisms, initial and long-range, is of particular interest. We examine briefly how this power surface depends on the two individual effects, on ξ_1 and on ξ_2 considered separately, and how it depends on their sum $\xi_1 + \xi_2$ which is an approximation to the total effect $\xi_1 + \xi_2 + \xi_1\xi_2$. When the two mechanisms are acting in the same direction, what is the power surface in a typical example, and how does this contrast with the power when the two mechanisms are acting in opposite directions? Is the asymptotic approximation for the power adequate with moderate sample sizes?

Neyman and Scott (1967c) investigated the power of the locally optimum $C(\alpha)$ test criterion Z_2 for detecting a change in the effect ξ_2 due to mechanism M_2 in a randomized experiment consisting of 100 independent trials, under the assumptions that the distribution is Gamma distributed with no predictor variables and that the treatment effect is multiplicative changing at most the scale parameter. The power functions of three nonparametric tests, the Wilcoxon, Mann-Whitney rank test, the Kolmogorov-Smirnov test and the median test, were studied at the same time since these tests are sometimes employed. The studies were made by Monte Carlo simulation for typical cases arising in weather modification experimentation, such as $\theta = 0.8$ for the untreated probability that precipitation will occur, and $\gamma = 0.6$, $\delta = 1.0$ as the untreated parameters in the Gamma distribution. With $n = 100$ experimental units, the power was discouragingly low for all

four tests. Even with level of significance 0.10, the probability of detecting a multiplicative effect of 1.5, corresponding to an increase of 50%, was slightly less than 0.6 for the locally optimum Z_2 test, and is lower still, about 0.45, for the Wilcoxon, Mann-Whitney test, and even smaller for the Kolmogorov-Smirnov and the median tests, a little more than 0.3 and a little less than 0.3, respectively. In these studies, the ordering of the power functions of these four tests was retained.

In the Monte Carlo studies reported here, we have continued comparisons with the Wilcoxon, Mann-Whitney test, labelled U and drawn with a dashed line in the figures. Figure 2 gives a comparison of the Monte Carlo power of the locally optimum Z_3 test criterion for testing the effect per experimental unit (solid line) as a function sum $\xi_1 + \xi_2$ with the asymptotic theoretical power (dotted line). In each panel the value of ξ_1 is fixed so that across a panel the value of ξ_2 is increasing, negative at the left of the panel and positive at the right, with the point of changeover through zero shifting as ξ_1 is increased. The case considered is similar to that in the earlier paper except that 200 experimental units are considered in the randomized trials since we now know that at least 200 trials are needed to achieve a reasonable experiment. The asymptotic power function provides a reasonable approximation for practical purposes except in those categories where ξ_1 is quite negative when the asymptotic power is too high especially when ξ_2 is large positive.

Figure 2 also shows the power function of the criterion Z_2 for comparison since, as noted above, some evaluations of the experiment have been made using only the positive observations. Unless ξ_1 is near zero (the center panel), the disagreement with the power functions of Z_3

EFFECT PER EXPERIMENTAL UNIT

CASE: $\gamma = 0.6$, $\delta = 1.0$, $\theta = 0.8$; No. Treat. + No. Not Treat = 200; No. Samples = 500; Level 0.10

.... Z_3 Asymptotic   ●——— Z_3,   △———— Z_2,   ○------ U Monte Carlo
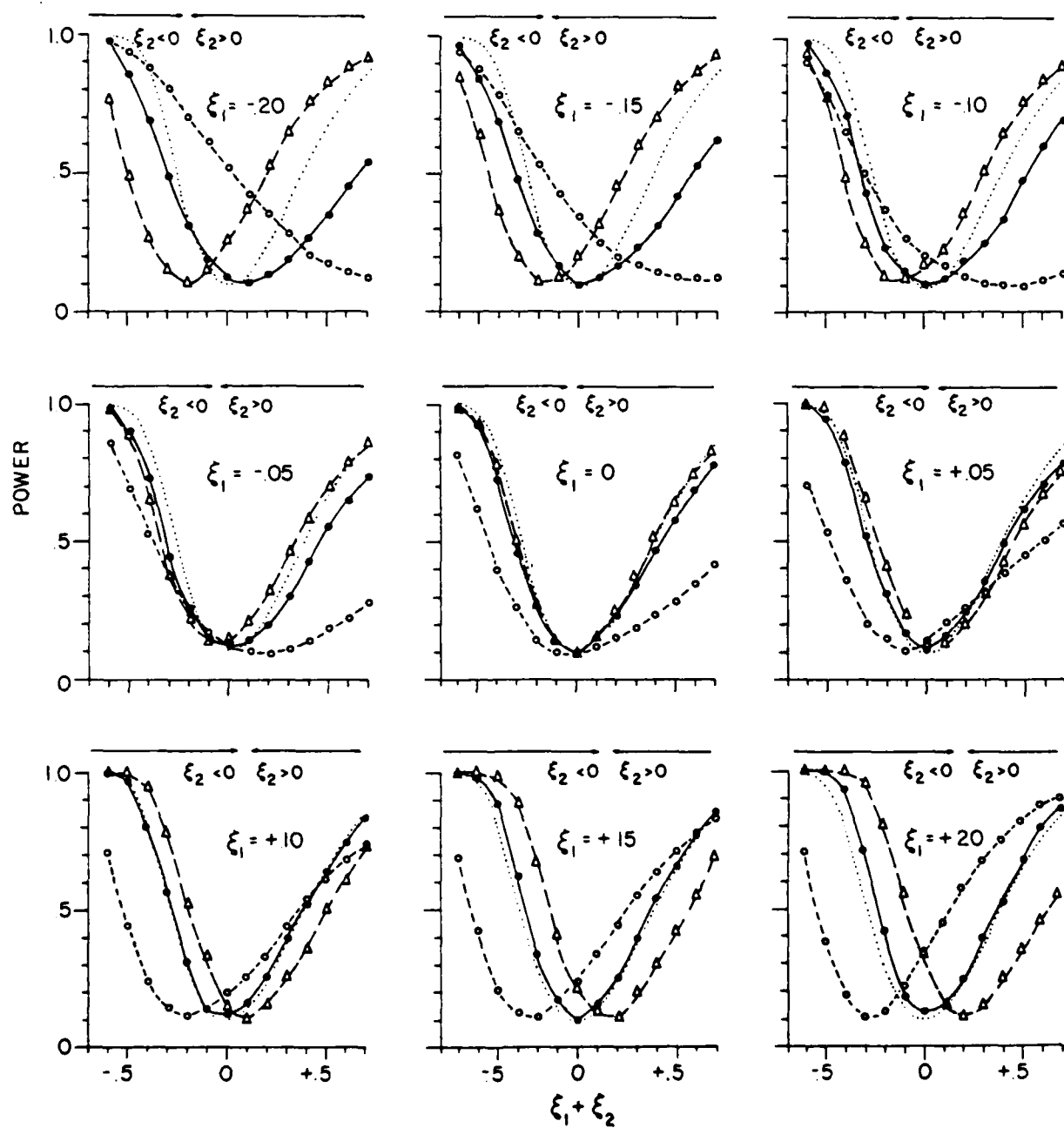

Figure 2

is pronounced. In particular, the power function of Z_2 has its minimum when ξ_2 is zero, not when $\xi_1 + \xi_2$ is zero so that, unless $\xi_1 = 0$, the use of Z_2 to test for combined effect produces a test bias which can be very large. When ξ_1 is negative, Z_2 has little chance of detecting that the combined effect is not zero when it is in fact negative, but has probability of several times the level of significance of finding that the effect is nonzero when it is actually zero. When the combined effect is positive, the power continues high. When ξ_1 is positive, the power function of the criterion Z_2 is a reflection of that just described. We thus conclude that when M_1 and M_2 are acting in the same direction, so that ξ_1 and ξ_2 have the same sign, the Z_2 test criterion has very little chance of detecting that the combined effect is not zero, even when the total of the two effects is quite large. However, when the two mechanisms are acting in opposite directions, the power of Z_2 is greater than that of Z_3. Unfortunately, this phenomenon persists even when the combined effect is zero, making the test invalid unless $\xi_1 = 0$ also.

The power function of the Wilcoxon, Mann-Whitney is even more bizarre. As indicated by the short-dashed lines, the test bias is large unless ξ_1 is near zero in which case the power function is much lower than that of competing tests. When ξ_1 and ξ_2 have opposite signs, the power of the U test tends to be very low, approximately the level of significance. When the mechanisms are in the same direction the power increases but this is not helpful since in just these categories the U test is very invalid, with a large probability of finding a nonzero effect when none exists.

We would like a method for estimating the effectiveness of the combined mechanism, or for testing that the effect is zero, that is more powerful than the test criterion Z_3. Several former colleagues in the Statistical Laboratory including Barry and Kang Ling James, S. Odoom, and Paul Wang are investigating these problems. Their studies are not yet completed and will be reported elsewhere.

Acknowledgement

References

DAWKINS, S.M., NEYMAN, J., SCOTT, E. L., AND WELLS, M. A. (1977) Preliminary statistical analysis of precipitation amounts--chapter 13. Data base and methodology of analysis. *The Final Report on the Pyramid Lake Pilot Project 1970 to 1975* (ed. P. Squires) Desert Research Institute, Reno, Nevada, 13.1-13.44.

MORAN, P. A. P. (1955) The statistical design of an experiment to test the stimulation of rain. *Australian J. Physics* 8, 440-448.

MORAN, P. A. P. (1973) Asymptotic properties of homogeneity tests. *Biometrika* 60, 79-85.

MOSTELLER, F. (1977) Experimentation and innovations. *Bull. Internat. Statist. Inst.* 47-1, 559-572.

NEYMAN, J. AND SCOTT, E.L.(1965) Asymptotically optimal tests of composite hypotheses for randomized experiments with noncontrolled predictor variables. *J. Amer. Statist. Assoc.* 60, 699-721,

NEYMAN, J. AND SCOTT, E. L. (1967a) Note on techniques of evaluation of single rain stimulation experiments. *Proc. 5th Berkeley Symp. Math. Statist. Prob.* 5, 371-384, Univ. California Press, Berkeley.

NEYMAN, J. AND SCOTT, E. L. (1967b) On the use of $C(\alpha)$ tests of composite hypotheses. *Bull. Internat. Statist. Inst.* 41-1, 477-496.

NEYMAN, J. AND SCOTT, E. L. (1967c) Some outstanding problems relating to
   rain modification. *Proc. 5th Berkeley Symp. Math. Statist. Prob.* 5,
   293-326, Univ. California Press, Berkeley.

NEYMAN, J., SCOTT, E. L., AND VASILEVSKIS, M. (1960) Statistical evaluation
   of the Santa Barbara randomized cloud seeding experiment. *Bull. Amer.*
   Meteorol. Soc. 41, 531-547.

NEYMAN, J., SCOTT, E. L., AND WELLS, M. A. (1968) Influence of atmospheric
   stability layers on the effect of ground-based cloud seeding, I. Empirical
   results. *Proc. Natl. Acad. Sci. USA* 60, 416-423.

SÄNGER, R. *et al.*(1958-1964) Grossversuch III zum Bekämpfung des Hagels
   im Tessin. *Annual Reports, 1957-1963.* Abteilung fur Landwirtschaft
   des Eidg. Volkswirtschaftsdepartmentes, Bern, Switzerland.

LEGENDS

Figure 1

Typical comparison of observed distribution of nonzero precipitation
with Gamma distribution fitted by maximum likelihood with same shape
parameter. These data correspond to the six stations with altitude
< 1000 km in zone 4 of the Swiss hail experiment Grossversuch III.

Figure 2

Power function for several tests that the combined effect per experimental
unit is zero. Comparison of the asymptotic theoretical power for Z_3 with
Monte Carlo simulated power for Z_3, for Z_2, and for Wilcoxon, Mann-Whitney
for fixed values of the initial effect ξ_1 and increasing values of the sum
of the two effects (and thus increasing values of the long-range effect ξ_2).